ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
UIT TOÀN DIỆN • SÁNG TẠO • PHỤNG SỰ

BÀI GIẢNG
XÁC SUẤT THỐNG KÊ

Giảng viên: TS. PHÙNG MINH ĐỨC

(Bộ môn Toán Lý)

Chương 4: Vectơ ngẫu nhiên

4.1 Vectơ ngẫu nhiên rời rạc

4.2 Vectơ ngẫu nhiên liên tục

4.3 Kỳ vọng, hiệp phương sai và hệ số tương quan

4.1 Vectơ ngẫu nhiên rời rạc

Định nghĩa 4.1.1

1. Cho X, Y là các biến ngẫu nhiên. Khi đó $Z = (X, Y)$ được gọi là một **vectơ ngẫu nhiên.**
2. Nếu X, Y là các biến ngẫu nhiên rời rạc (liên tục) thì (X, Y) được gọi là **vectơ ngẫu nhiên rời rạc (liên tục)**.

Định nghĩa 4.1.2

1. Cho X, Y là các biến ngẫu nhiên rời rạc. Xác suất đồng thời của X, Y (joint probability) được xác định bởi

$$P(x, y) = P(X = x, Y = y).$$

Bảng phân phối xác suất đồng thời

$X \backslash Y$	y_1	y_2	$\cdots$	y_m
x_1	p_{11}	p_{12}	$\cdots$	p_{1m}
x_2	p_{21}	p_{22}	$\cdots$	p_{2m}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nm}

trong đó $p_{ij} = P(X = x_i, Y = y_j)$.

Định nghĩa 4.1.2

2. Phân phối của từng biến X, Y được gọi là **phân phối xác suất thành phần** (phân phối biên) (marginal probability distributions) được xác định như sau

$$P_X(x_i) = P(X = x_i) = \sum_{j=1}^{m} P(x_i, y_j)$$

$$P_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{n} P(x_i, y_j)$$

Bảng phân phối xác suất của X

X	x_1	x_2	$\cdots$	x_n
$P(X = x_i)$	p_{1*}	p_{2*}	$\cdots$	p_{n*}

trong đó $p_{i*} = p_{i1} + p_{i2} + \cdots + p_{im}$.
Kỳ vọng (trung bình) thành phần của X

$$E(X) = \sum_{i=1}^{n} x_i p_{i*}$$

Bảng phân phối xác suất của Y

Y	y_1	y_2	$\cdots$	y_m
$P(Y = y_j)$	p_{*1}	p_{*2}	$\cdots$	p_{*m}

trong đó $p_{*j} = p_{1j} + p_{2j} + \cdots + p_{nj}$.
Kỳ vọng (trung bình) thành phần của Y

$$E(Y) = \sum_{j=1}^{m} y_j p_{*j}.$$

Định nghĩa 4.1.3 (Phân phối xác suất có điều kiện)

Cho X, Y là các biến ngẫu nhiên rời rạc. Xác suất của X khi đã biết $Y = y_j, P_Y(y_j) > 0$:

$$P(X = x_i | Y = y_j) = \frac{P(x_i, y_j)}{P_Y(y_j)} = \frac{p_{ij}}{p_{*j}}.$$

Xác suất của Y khi đã biết $X = x_i, P_X(x_i) > 0$:

$$P(Y = y_j | X = x_i) = \frac{P(x_i, y_j)}{P_X(x_i)} = \frac{p_{ij}}{p_{i*}}.$$

Bảng phân phối xác suất của X với điều kiện $Y = y_j$

X	x_1	x_2	$\cdots$	x_n
$P(X = x_i \vert Y = y_j)$	$\frac{p_{1j}}{p_{*j}}$	$\frac{p_{2j}}{p_{*j}}$	$\cdots$	$\frac{p_{nj}}{p_{*j}}$

Kỳ vọng của X với điều kiện $Y = y_j$

$$E(X|Y = y_j) = x_1 \frac{p_{1j}}{p_{*j}} + x_2 \frac{p_{2j}}{p_{*j}} + \cdots + x_n \frac{p_{nj}}{p_{*j}}.$$

(tương tự với bảng phân phối XS và kỳ vọng của Y khi đã biết $X = x_i$)

Định nghĩa 3.1.3

Hai biến ngẫu nhiên rời rạc X, Y là **độc lập** nếu

$$P(x, y) = P_X(x)P_Y(y), \forall x, y.$$

Ví dụ 4.1.1 Một chương trình bao gồm hai mô-đun. Đặt X là số lỗi trong mô-đun 1 và Y là số lỗi trong mô-đun 2 có xác suất đồng thời như sau $P(0,0) = P(0,1) = P(1,0) = 0,2; P(1,1) = P(1,2) = P(1,3) = 0,1; P(0,2) = P(0,3) = 0,05.$

a. Tìm phân phối xác suất thành phần của X.

b. Tìm phân phối của tổng số lỗi trong chương trình.

c. Các lỗi trong hai mô-đun có xảy ra độc lập hay không?

d. Giả sử chương trình có lỗi. Tính xác suất mô-đun 1 có lỗi.

e. Giả sử mô-đun 1 có lỗi. Tính xác suất mô-đun 2 có lỗi.

Giải Bảng phân phối xác suất đồng thời của X, Y như sau

X \ Y	0	1	2	3
0	0,2	0,2	0,05	0,05
1	0,2	0,1	0,1	0,1

a.
b. $Z = X + Y \in \{0, 1, 2, 3, 4\}$
Tính $P(Z = k) = \cdots, k = 0, 1, 2, 3, 4$
c.
d. Tính $P(X \geq 1 | Z \geq 1) = \cdots$
e. Tính $P(X \geq 1 | Y \geq 1) = \cdots$

4.2 Vectơ ngẫu nhiên liên tục

Định nghĩa 4.2.1

Cho X, Y là các biến ngẫu nhiên liên tục.

1. **Hàm mật độ xác suất đồng thời** (joint probability density function) của hai biến ngẫu nhiên là một hàm $f(x, y) \geq 0$ thỏa mãn
$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \mathrm{d}x \mathrm{d}y = 1.$$
2. Hàm mật độ xác suất thành phần (marginal probability density function) của X và Y được lần lượt xác định như sau
$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy; \; f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Định nghĩa 4.2.1

3. **Hàm phân phối xác suất đồng thời** (joint probability density function) của hai biến ngẫu nhiên

$$F(x,y) = P(X \le x, Y \le y).$$

4. Hai BNN X, Y gọi là độc lập nếu

$$P(X \le x, Y \le y) = P(X \le x).P(Y \le y)$$
$$\Leftrightarrow F(x,y) = F_X(x).F_Y(y), \quad \forall x, y.$$

Mệnh đề 4.2.1

Cho X, Y là các biến ngẫu nhiên liên tục. Các điều sau là tương đương

1. X, Y là độc lập
2. $f(x,y) = f_X(x).f_Y(y), \forall x, y$

Mệnh đề 4.2.2

Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời $f(x,y)$. Khi đó

1. $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y) \mathrm{d}x \mathrm{d}y$
2. $F(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^b \int_{-\infty}^a f(x,y) \mathrm{d}x \mathrm{d}y$

   Tổng quát: $P[(X,Y) \in D] = \iint_D f(x,y) \mathrm{d}x \mathrm{d}y$
3. $f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y).$

Ví dụ 4.2.1 Hàm mật độ xác suất đồng thời của các biến ngẫu nhiên X và Y được cho như sau

$$f(x,y)=\begin{cases} Ce^{-x}e^{-2y}, & x>0, y>0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm C.
b. Tìm hàm mật độ thành phần của X, Y. Hai biến X và Y có độc lập không?
c. Tính $P(X>1, Y<1)$ và $P(X>1|Y<1)$.
d. Tính $P(X<Y)$.

Giải.

a. $1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y = \cdots = \int_{0}^{+\infty}\int_{0}^{+\infty} e^{-x}e^{-2y}\mathrm{d}x\mathrm{d}y =$
$\cdots \Rightarrow C = 2.$

b. $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy:$

- $x \le 0:$
- $x > 0:$

c. $P(X > 1, Y < 1) = \cdots = \int_{1}^{+\infty}\mathrm{d}x\int_{0}^{1} e^{-x}e^{-2y}\mathrm{d}y = \cdots$

$P(Y < 1) = \cdots = \int_{0}^{+\infty}\mathrm{d}x\int_{0}^{1} e^{-x}e^{-2y}\mathrm{d}y = \cdots$

$\Rightarrow P(X > 1|Y < 1) = \dfrac{P(X > 1, Y < 1)}{P(Y < 1)} = \cdots$

d. $P(X < Y) = \int_{-\infty}^{y} f_X(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{y} f(x,y)dxdy = .$

Định nghĩa 4.2.2

1. Cho X, Y là các biến ngẫu nhiên liên tục. **Hàm mật độ xác suất có điều kiện** của X khi đã biết $Y = y_0$

$$f_X(x|Y = y_0) = \begin{cases} \dfrac{f(x, y_0)}{f_Y(y_0)}, & f_Y(y_0) > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

2. **Xác suất có điều kiện** của X khi đã biết $Y = y_0$

$$P(a < X < b|Y = y_0) = \int_a^b f_X(x|Y = y_0)dx.$$

Phát biểu tương tự bằng cách đổi vai trò của X và Y cho hàm mật độ và xác suất có điều kiện của Y khi biết $X = x_0$.

Chú ý:

Khi X, Y là các biến ngẫu nhiên liên tục thì $P(Y = y_0) = 0$ nên **không** tính Xác suất có điều kiện của X khi đã biết $Y = y_0$ theo công thức xác suất có điều kiện

$$P(a < X < b | Y = y_0) = \frac{P(a < X < b, Y = y_0)}{P(Y = y_0)}.$$

Ví dụ 4.2.2 Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau

$$f(x,y) = \begin{cases} \frac{2}{5}(2x + 3y) & , 0 \le x \le 1, 0 \le y \le 1 \\ 0 & , \text{ các trường hợp còn lại} \end{cases}$$

a. Tìm hàm mật độ có điều kiện $f_Y(y|x)$.

b. Tính $P(\frac{1}{4} < Y < 1 | X = \frac{3}{4})$.

Giải. a. Hàm mật độ thành phần của X là

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \ldots$$

Tìm hàm mật độ điều kiện $f_Y(y|x)$:

$$f_Y(y|x) = \begin{cases} \dfrac{f(x,y)}{f_X(x)}, & 0 \le x, y \le 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

b. Tính

$$P(\frac{1}{4} < Y < 1|X = \frac{3}{4}) = \int_{\frac{1}{4}}^{1} f_Y(y|X = \frac{3}{4})dy = \ldots$$

Bài tập 1. Cho X, Y là các biến ngẫu nhiên và hàm số

$$f(x,y) = \begin{cases} 10x^2y, & 0 \le y \le x \le 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Chứng tỏ rằng (X, Y) có hàm mật độ đồng thời là $f(x,y)$.

b. Tính $P(Y \ge \frac{1}{2}X)$.

c. Tìm hàm mật độ thành phần của X, Y.

d. Tìm hàm mật độ có điều kiện $f_X(x|y), f_Y(y|x)$.

e. Tính $P(Y < \frac{1}{8}|X = \frac{1}{4})$.

4.3 Kỳ vọng, hiệp phương sai và hệ số tương quan

Mệnh đề 4.3.1

Cho X, Y là các biến ngẫu nhiên và một hàm $h(X, Y)$. Kỳ vọng của hàm $h(X, Y)$, ký hiệu là $E(h(X, Y))$, được xác định như sau

1. Nếu X, Y là các biến ngẫu nhiên rời rạc thì

$$E(h(X,Y)) = \sum_x \sum_y h(x,y)P(x,y)$$

2. Nếu X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời $f(x, y)$ thì

$$E(h(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy$$

Ví dụ 4.3.1

Cho X, Y là các biến ngẫu nhiên liên tục có hàm mật độ đồng thời xác định như sau

$$f(x,y)=\begin{cases} 24xy, & 0\le x\le 1; 0\le y\le 1; x+y\le 1\\ 0, & \text{các trường hợp khác}\end{cases}$$

và hàm $h(X,Y)=0,75+0,75X+1,5Y$. Tính $E(h(X,Y))$.

Giải. Ta có

$$\begin{aligned} E(h(X,Y)) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy\\ &= \int_0^1\int_0^{1-y}(0,75+0,75x+1,5y)24xydxdy\\ &= 1,65.\end{aligned}$$

Định nghĩa 4.3.1

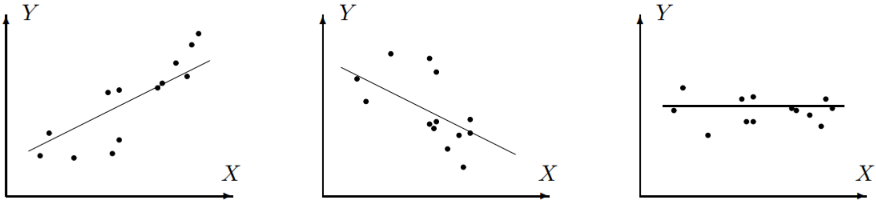
Cho X, Y là các biến ngẫu nhiên.
1. **Hiệp phương sai** (covariance) của X và Y, ký hiệu $\mathrm{Cov}(X, Y)$, được xác định bởi

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

2. **Hệ số tương quan** (Correlation coefficient) của X, Y được xác định như sau

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

- Nếu $\operatorname{Cov}(X, Y) > 0$ thì X tăng suy ra Y tăng và X giảm suy ra Y giảm.
- Nếu $\operatorname{Cov}(X, Y) < 0$ thì X tăng suy ra Y giảm và X giảm suy ra Y tăng.
- Nếu $\operatorname{Cov}(X, Y) = 0$ thì ta nói X, Y không *tương quan*.



(a) $\operatorname{Cov}(X, Y) > 0$ (b) $\operatorname{Cov}(X, Y) < 0$ (c) $\operatorname{Cov}(X, Y) = 0$

- Nếu $|\rho| = 1$ thì ta nói các điểm (x_i, y_j) nằm trên một đường thẳng.
- Nếu ρ gần 1 thì ta nói X, Y có tương quan dương mạnh.
- Nếu ρ gần -1 thì ta nói X, Y có tương quan âm mạnh.
- Nếu ρ gần 0 thì ta nói X, Y có tương quan yếu hoặc không tương quan.

Ví dụ 4.3.2 : xem ví dụ 4.1.1

Ví dụ 4.3.3 Cho X, Y là các biến ngẫu nhiên liên tục và hàm mật độ xác suất đồng thời

$$f(x,y) = \begin{cases} 2, & x + y \leq 1, x > 0, y > 0 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Tính $\mathrm{Cov}(X, Y)$ và $\rho(X, Y)$.

Giải.

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = \int_0^1\int_0^{1-y} 2xydxdy$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$E(Y) = \ldots\ldots\ldots\ldots\ldots$$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) =$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx$$

$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$E(Y^2) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$V(X) = E(X^2) - (EX)^2 =$$

$$V(Y) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} =$$

Mệnh đề 4.3.2

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
2. $\rho(X, Y) = \rho(Y, X)$
3. $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$
4. $\rho(aX + b, cY + d) = \rho(X, Y)$
5. Nếu X, Y độc lập thì $\mathrm{Cov}(X, Y) = 0.$

Bài tập 2. Cho hàm mật độ xác suất đồng thời của X, Y như sau

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1, 0 < y < 1 - x \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tính trung bình thành phần của X, Y.
b. Tính $P(X > 0,3 | Y = 0,5)$.
c. X, Y có độc lập không?
d. Tính $P(X + Y \leq 0,5), P(Y \geq 0,5)$.

Bài tập 3. Cho hàm mật độ xác suất đồng thời của X, Y như sau

$$f(x,y) = \begin{cases} Cx^2y, & 0 \le y \le x \le 1 \\ 0, & \text{các trường hợp khác} \end{cases}$$

a. Tìm C.

b. Tính $P(Y \ge \frac{1}{2}X)$.

c. Tìm hàm mật độ thành phần của X, Y.

d. Tính trung bình thành phần của X, Y.

e. X, Y có độc lập không?

f. Tìm hàm mật độ có điều kiện $f_X(x|y)$.

g. Tính $P(Y > 1|X = \frac{1}{3})$.

h. Tính $P(X > \frac{1}{3}|Y > \frac{1}{4})$.